



**UNIVERSITÉ DE
MONTPELLIER**

Lecture d'article

ADAM pour le Deep Learning

Groupe :

ACHIQ Aya

CLETZ Laura

EL MAZZOUJI Wahel

2 Juin 2025 - 25 Juillet 2025

Table des matières

1 Introduction

Contexte : importance des méthodes d'optimisation stochastique. Problématique : Adam est populaire car il converge rapidement, mais sa capacité de généralisation reste discutée. Présenter Adam et voir ses limites, grâce à une illustration numérique. Est ce que cette méthode est meilleure que SGD ?

2 La méthode Adam (Kingma Ba, 2014)

2.1 Idée générale

Adam combine le momentum et une adaptation du taux d'apprentissage inspirée de Ada-Grad/RMSProp.

2.2 Algorithme

Expliquer ce que fait l'algorithme, sans rentrer dans les détails (montrer le pseudo code le jour de l'oral.)

2.3 Forces

— Convergence rapide et stable. — Hyperparamètres par défaut robustes ($\beta_1 = 0.001$, $\beta_2 = 0.9$, $\epsilon = 10^{-8}$). — Performances solides sur des réseaux de neurones convolutifs et bien d'autres

3 Limites d'Adam et méthodes adaptatives

3.1 Résultats théoriques

Sur des problèmes surparamétrés, Adam peut converger vers des solutions qui généralisent mal, contrairement à SGD qui maximise la marge.

3.2 Résultats empiriques

Voir les résultats sur les expériences sur CIFAR-10, War Peace. Idée : Adam cvg vite mais donne de moins bons résultats en test.

3.3 Implications pratiques

— SGD reste préférable quand la généralisation est importante. — Adam utile pour du reinforcement learning par exemple.

3.4 Illustration numérique

Comparaison Adam vs SGD ? On affiche les résultats mais pas les codes (les mettre en annexe à voir)

4 Conclusion

On termine par une conclusion, on dit quelques mots sur ADAM (il est énormément utilisé, efficace etc...) mais possède des défauts. On fait une ouverture sur des méthodes plus récentes peut etre.

Annexes

Ici, on mettra le code et peut etre l'algorithme d'ADAM